

Corrigé des exercices

Exercice 1 :

On cherche la loi du couple (U,V), on a :

$$E(U) = E(X + Y + Z) = E(X) + E(Y) + E(Z) = 0$$

$$E(V) = E(X - Y) = E(X) - E(Y) = 0$$

et d'autre part,

$$\text{var}(U) = \text{var}(X + Y + Z) = \text{var}(X) + \text{var}(Y) + \text{var}(Z) = 3$$

$$\text{var}(V) = \text{var}(X - Y) = E((X - Y)^2) - E(X - Y)^2 = \text{var}(X) + \text{var}(Y) = 2$$

et

$$\begin{aligned} \text{cov}(U, V) &= E(UV) - E(U)E(V) = E((X + Y + Z)(X - Y)) \\ &= E(X^2) - E(Y^2) + E(XZ) - E(YZ) = \text{var}(X) - \text{var}(Y) = 0 \end{aligned}$$

donc

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \right)$$

Comme cette loi est normale alors comme les covariances sont nulles alors les variables U et V sont indépendantes.

Exercice 2 :

La combinaison linéaire d'une variable aléatoire multinormale est multinormale.

On a $E(Y) = E(AZ + c) = E(AZ) + c = AE(Z) + c$.

Pour la variance,

$$\begin{aligned} \text{var}(Y) &= \text{var}(AZ + c) = \text{var}(AZ) \\ &= E((AZ)^2) = E(AZZ^T A^T) \\ &= AE(ZZ^T)A^T = A \text{var}(Z)A^T \\ &= AI_p A^T = AA^T \end{aligned}$$

Exercice 3 :

Afin de trouver l'ellipsoïde d'isodensité, on utilise la distance de Mahalanobis, celle-ci suit une loi du χ_p^2 à p degré de liberté (p étant la dimension du vecteur traité, ici p=2). On calcule donc dans le premier cas :

$$D^2 = \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + \frac{y^2}{4}$$

Comme $D^2 \sim \chi_2^2$, on veut $P(\mathbf{x} \in E_k) = 0,95 \Rightarrow P(x < \chi_2^2) = P(x < 6) = 0,95$, le k vaudra 6, on veut donc résoudre :

$$x^2 + \frac{y^2}{4} = 6 \Leftrightarrow \left(\frac{x}{\sqrt{6}}\right)^2 + \left(\frac{y}{2\sqrt{6}}\right)^2 = 1$$

On a $\sqrt{6} \approx 2,45$. Une ellipse de centre (0,0) et de rayon 2,45 sur les x et 4,9 sur les y.

De même pour P_2 .

$$\text{On obtient } \frac{(x-2)^2}{4} + (y-2)^2 = 6 \Leftrightarrow \left(\frac{x-2}{2\sqrt{6}}\right)^2 + \left(\frac{y-2}{\sqrt{6}}\right)^2 = 1$$

Une ellipsoïde de centre (2,2) et de rayon 4,9 sur les x et 2,45 sur les y.

Exercice 4 :

1) On a $\hat{\mu} \sim N(\mu, \sigma^2/n)$, on aura donc $\pi = \frac{\hat{\mu} - \mu}{\sqrt{\sigma^2/n}}$, on utilise

$$P(-u_{1-\alpha/2} \leq \pi \leq u_{1-\alpha/2}) = 1 - \alpha$$

$$\text{On obtient un IC : } \left[\hat{\mu} - 1,96\sqrt{\frac{\sigma^2}{n}}; \hat{\mu} + 1,96\sqrt{\frac{\sigma^2}{n}} \right]$$

Pour l'intervalle de tolérance, on utilise la formule du cours. Dans le cas univarié, on connaît la variance, on trouve une transformation afin d'avoir une loi standard.

On a

$$x - g \sim N\left(0, \sigma^2\left(1 + \frac{1}{n}\right)\right)$$

$$\frac{x - g}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}} \sim N(0,1)$$

On a donc

$$P \left(-u_{1-\alpha/2} \leq \frac{x-g}{\sqrt{\sigma^2 \left(1 + \frac{1}{n}\right)}} \leq u_{1-\alpha/2} \right) = 1 - \alpha$$

L'intervalle de tolérance est tel que :

$$IC : \left[\bar{X} - u_{1-\alpha/2} \sqrt{\sigma^2 \left(1 + \frac{1}{n}\right)}; \bar{X} + u_{1-\alpha/2} \sqrt{\sigma^2 \left(1 + \frac{1}{n}\right)} \right]$$

2) Si la variance est inconnue, on utilise l'estimation de celle-ci :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2, \text{ on a que}$$

$$\frac{\hat{\mu} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n-1}}} \sim T_{n-1}$$

La loi de Student ne dépend ni de μ ni de la variance.

On a donc un intervalle de confiance IC :

$$P \left(-t_{n-1}(\alpha/2) \leq \frac{\hat{\mu} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n-1}}} \leq t_{n-1}(\alpha/2) \right) = 1 - \alpha$$

$$IC : \left[\bar{X} - t_{n-1}(\alpha/2) \sqrt{\frac{\hat{\sigma}^2}{n-1}}; \bar{X} + t_{n-1}(\alpha/2) \sqrt{\frac{\hat{\sigma}^2}{n-1}} \right]$$

et un intervalle de tolérance :

$$P \left(-f_{1,n-1}(\alpha/2) \leq \frac{n}{n+1} \frac{(x-g)^2}{\hat{\sigma}^2} \leq f_{1,n-1}(\alpha/2) \right) = 1 - \alpha$$

$$\left[g - \sqrt{\frac{n+1}{n} f_{1,n-1}(\alpha/2) \hat{\sigma}^2}; g + \sqrt{\frac{n+1}{n} f_{1,n-1}(\alpha/2) \hat{\sigma}^2} \right]$$

Exercice 5 :

Pour l'ellipsoïde de confiance de la moyenne, on utilise les formules du cours et on obtient donc :

$$\begin{aligned} \begin{pmatrix} 2-\mu_1 \\ 2-\mu_2 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 2-\mu_1 \\ 2-\mu_2 \end{pmatrix} &= \frac{1}{50-2} F(2; 50-2) \\ \Leftrightarrow (2-\mu_1)^2 + \frac{(2-\mu_2)^2}{4} &= \frac{4}{48} = \frac{1}{6} \\ \Leftrightarrow \left(\frac{2-\mu_1}{1/\sqrt{6}} \right)^2 + \left(\frac{2-\mu_2}{\sqrt{2}/\sqrt{3}} \right)^2 &= 1 \end{aligned}$$

Ellipsoïde de centre (2,2) et de rayon 0,408 sur les x et 0,816 sur les y.

Pour l'intervalle de tolérance, on a donc :

$$\begin{aligned} \begin{pmatrix} x-2 \\ y-2 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} x-2 \\ y-2 \end{pmatrix} &= \frac{(50-1)2}{50-2} \frac{50+1}{50} F(2; 50-2) \\ \Leftrightarrow (x-2)^2 + \frac{(y-2)^2}{4} &= \frac{4998}{2400} 4 \approx 8.33 \\ \Leftrightarrow \left(\frac{x-2}{\sqrt{8.33}} \right)^2 + \left(\frac{y-2}{2\sqrt{8.33}} \right)^2 &= 1 \end{aligned}$$

Ellipsoïde de centre (2,2) et de rayon 2,88 sur les x et 5,77 sur les y.

Exercice 6 :

On teste l'égalité de deux moyennes avec le logiciel SAS et la proc discrim que vous étudierez plus tard dans le cours.

Pages 149 – 150 : les données,

p. 163, stat générales

p. 155, matrices de covariances intra-classes (pour chaque classe)

p. 156, moyenne des matrices de covariances intra-classes calculée par :

$$W = \frac{1}{n} \sum_{k=1}^2 n_k V_k$$

et matrice inter-classes, matrice de covariances obtenue à partir des centres de gravité de chaque classe

p. 157, matrice de covariance de l'ensemble de l'échantillon, distance de Mahalanobis entre les classes, calculé par :

$$D_{ij}^2 = (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}_j)^T V^{-1} (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}_j)$$

On a $D^2 = 4,94 \neq 0$. Si on crée un intervalle de confiance autour par cette distance. On a

$$\frac{n_1 n_2}{n} \frac{n-p-1}{p(n-2)} D_p^2 = \frac{50 \cdot 49}{99} \frac{99-7-1}{7(99-2)} 4,942 = 16,39$$

et $F(7,91) = 2,12$

On n'a donc pas égalité entre les moyennes.

Exercice 7 :

1- La période est de 3.

2- La densité de la loi de Cauchy est donnée par $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$. Sa fonction de répartition est donc donnée par : $F(t) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(t)$ (on ajoute $\frac{1}{2}$ afin de bien définir l'intervalle sur lequel la fonction est définie : $[-\pi/2; \pi/2]$). On utilise la méthode d'inversion, on obtient donc pour une variable aléatoire R de distribution uniforme sur $[0; 1]$:

$$x = F^{-1}(u) = \tan\left(\pi\left(u - \frac{1}{2}\right)\right).$$

Il suffit de simuler une variable uniforme et d'utiliser la transformation ci-dessus.

- 3- La loi de Poisson peut être reliée à la loi exponentielle. En effet, si des évènements surviennent à des dates séparées par des durées exponentielles de paramètre λ , le nombre d'évènements survenant en une unité de temps suit une loi de Poisson de même paramètre.

On simule des variables aléatoires $\{Y_1, Y_2, \dots\}$ i.i.d. $Y_i \sim \exp(\lambda)$ à partir de la loi uniforme en utilisant la fonction de répartition inverse et on définit :

$$X = \sum_{k \geq 0} k \cdot 1\{Z_k \leq 1 \leq Z_{k+1}\}$$

avec $Z_k = \sum_{i=1}^k Y_i$

Il faut donc simuler des variables aléatoires exponentielles de paramètre λ et compter le nombre de simulations nécessaires pour dépasser 1, ou bien simuler des variables aléatoires exponentielles de paramètre 1 et compter le nombre de simulations nécessaires pour dépasser λ .

Exercice 8 :

- 1- On utilisera comme bornes 5% et 95%.
- 2- Algorithme à utiliser :
 - a. Tirer aléatoirement B échantillons de taille n avec remise dans (X_1, \dots, X_n) . On tire donc des échantillons de distribution $F_n(X)$.
 - b. Calculer la statistique T^b pour chaque échantillon ($b=1, \dots, B$).
 - c. La distribution d'échantillonnage $G(x)$ de T peut être approximé par la distribution empirique des des T^b :

$$G(x) \approx \frac{1}{B} \sum_{b=1}^B I(T^b \leq x)$$

- d. Le biais, la variance et l'intervalle de confiance à 95% associés à T peuvent être approximés par :

$$Biais(T) \approx \frac{1}{B} \sum_{b=1}^B T^b - \theta(F_n)$$

avec $\theta(F_n)$ paramètre calculé pour la distribution empirique.

$$Variance(T) \approx \frac{1}{B} \sum_{b=1}^B (T^b)^2 - \left(\frac{1}{B} \sum_{b=1}^B T^b \right)^2$$

Recherche des percentiles $\alpha/2$ et $1-\alpha/2$ de $G(x)$. L'intervalle de confiance de type percentile de niveau $1-\alpha$ est alors défini par :

$$IC(T) = \left[v \cdot \frac{\alpha}{2}, v \cdot \left(1 - \frac{\alpha}{2}\right) \right]$$